Markov Decision Processes Lecture Notes 00

Course Information

Authors: Benjamin Van Roy and Yifan Zhu *September 22, 2025*

1 Purpose

Markov decision processes serve as a general framework for modeling sequential decision under uncertainty. The subject provides a foundation for operations research, artificial intelligence, communications, economics, and other fields. The increasing scale of data, computation, and automation are driving innovative applications that build on these foundations. This course offers an introduction to the subject that prepares students to develop deeper knowledge and contribute to these foundations and applications. After taking this course, students should be proficient in formulating Markov decision processes and decision objectives and developing algorithms and code to produce solutions.

2 Syllabus

The content of this course has evolved over decades of teaching at Stanford. Some of this material was developed by or together with Pete Veinott, Stephen Boyd, Sanjay Lall, Zheng Wen, and Morteza Ibrahimi. The course will cover:

1. MDP specification
   (a) transition probabilities
   (b) transition function and disturbance distribution
   (c) semi-MDP specification
2. MDP objectives
   (a) total reward
   (b) discounted reward
   (c) average reward; gain and bias
3. examples
   (a) artificial intelligence: robots, LLMs
   (b) operations research: queues, inventory, portfolios, pricing, bandits
4. general algorithms
   (a) value iteration
   (b) policy iteration
   (c) linear programming
   (d) Q-learning
   (e) policy gradient methods
5. specialized algorithms
   (a) linear-quadratic control
   (b) inventory management
   (c) portfolio management
   (d) mult-armed bandits
6. reward learning

The treatment of these topics will focus on modeling, algorithms, mathematical concepts, and intuition rather than formal mathematical proofs.

3 Prerequisites and Programs

The course will rely on knowledge of probability and programming. Requisite background in probability can be obtained through MS&E 121, EE 178, or CS 109, and for programming, CS106B. Students are expected, for example, to be proficient in working with Gaussian, exponential, and beta distributions, Markov chains, and notions of convergence. Students are expected to also have experience in developing and debugging software implementations of numerical algorithms.

This course satisfies requirements for the Data and Decisions and the Computational Social Sciences programs in MS&E and for the Information Systems and Science program in EE.

4 Administration

Lectures will take place Mondays and Wednesdays at 3:00-4:20PM in room 60-109. All course materials will be available through Canvas. Ben Van Roy's office hours will take place Fridays 3:00-4:00PM in Packard 273. Yifan Zhu will serve as CA and hold office hours Thursdays at 4:00-5:30PM in Packard 107. Emails for the teaching staff should be sent to mdpclassfall2025staff@lists.stanford.edu.

5 Assignments and Grading

Homework assignments will require a combination of analytic and computational work. Computational exercises will be carried out in Python on Colab. While students are expected to have prior experience with Python and Colab or similar development environments, this tutorial may serve as a useful review [Johnson et al.].

Homeworks will generally be due on Fridays, submitted and graded via Gradescope. The cutoff for on-time submission is midnight on the due date. Late days are counted in 24-hour periods. Submitting between midnight on the due date and midnight the next day is one day late, and so on. Late work is penalized 10% per day late. Homework will not be accepted if more than two days late.

Students are allowed to collaborate on homework assignments. However, each student must write up the solutions on their own. Generative AI tools are permitted on the homework assignments, but will not be available for the exams.

A mid-term exam will be administered in class on Wednesday, October 29. There will be a final exam administered during the university-scheduled time.

Course grades are based 25% on homeworks, 25% on the mid-term exam, and 50% on the final exam.

6 Course Material

Course content will be covered in lectures, and lecture notes will be available for reading to reinforce that content. More comprehensive treatments of the subject are provided by Bertsekas [2012, 2011], Puterman [2014]; these are not required for the course but serve as useful resources for further study.

References

Dimitri Bertsekas. *Dynamic programming and optimal control*, volume 2. Athena scientific, 2011.

Dimitri Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific, 2012.

Justin Johnson, Volodymyr Kuleshov, and Isaac Caswell. CS231n Python tutorial with Google Colab. URL `https://colab.research.google.com/github/cs231n/cs231n.github.io/blob/master/python-colab.ipynb`.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.